DOCUMENT RESUME

ED 263 876 IR 011 860

AUTHOR Clark, Richard E.; Leonard, Stuart
TITLE Computer Research Confounding.
SPONS AGENCY University of Southern California, Los Angeles.
PUB DATE 31 Mar 85
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985). Research made possible by a grant from the EDUCARE Faculty Research Projects Committee of the University of Southern California.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Computer Assisted Instruction; Conventional Instruction; Drills (Practice); Instructional Improvement; Instructional Innovation; Intermode Differences; Meta Analysis; Research Design; *Research Needs; *Research Problems; Tables (Data); Teaching Methods; *Validity
IDENTIFIERS *Instructional Effectiveness

ABSTRACT

The suspected sources of confounding in current meta-analytic studies of computer based instruction (CBI) are uncontrolled effects of instructional method and/or the John Henry Effect (i.e., compensatory rivalry). To determine which confounding is most plausible, a random 30% sample of the 128 studies which formed the original Kulik meta-analyses for primary, secondary, and college levels was selected for a descriptive review. Evidence from reviewer coding of study features strongly suggests that the achievement gains found in CBI studies are actually due to the uncontrolled but robust instructional methods employed in these treatments. Cautions are offered for those who wish to implement CBI in schools and for research design, and it is suggested that the achievement gains attributed to the computer mode of delivery are probably due instead to the instructional methods employed in their software, and these methods could probably be delivered by a teacher. A list of references and appendices with data tables complete the document. (Author/JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED263876

COMPUTER RESEARCH CONFOUNDING

Richard E. Clark
University of Southern California

Stuart Leonard
Xerox Corporation

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard E. Clark

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented to the annual convention of the American Educational Research Association, Chicago Illinois, March 31, 1985.

This research was made possible, in part, by a grant from the EDUCARE Faculty Research Projects Committee at the University of Southern California to the first author. Both authors gratefully acknowledge the support of the members of EDUCARE.

IR011860

ABSTRACT

The suspected sources of confounding in current meta-analytic studies of computer based instruction (CBI) are uncontrolled effects of instructional method and/or the John Henry Effect (i.e. compensatory rivalry). To determine which confounding is most plausible, a random 30% sample of the 128 studies which formed the original Kulik meta-analyses for primary, secondary and college were selected for a descriptive review. Evidence from reviewer coding of study features strongly suggest that the achievement gains found in CBI studies are actually due to the uncontrolled but robust instructional methods employed in these treatments. Cautions are offered for those who wish to implement CBI in schools and for research design.

COMPUTERS IN INSTRUCTION

There is considerable enthusiasm for using computer-based instruction (CBI) to remedy deficits instruction at all levels of education. While estimates of the scope of CBI applications vary, recent surveys indicate that 62% of the school teachers in the United States have used computers for instruction (Riccobono, 1984), primarily for science, mathematics and computer programming instruction. This is a larger percentage than the 51% who used instructional television last year. The hope that computers might play a significant role in future science instruction is enhanced by their penetration into schools. As of 1983, approximately 75% (55,000) of the school buildings in the Nation had a total of over 630,000 computers available for instructional use (Quality Educational Data, 1983-84). One strongly suspects that the number is considerably higher today. This suggests a more rapid acceptance by the teaching profession and a greater density of equipment penetration than with previous instructional media (Clark, In press).

Computer Effects on Learning. It is possible that the major source of support for CBI use in education is the overwhelmingly positive results of research comparing CBI with more traditional teaching methods (e.g. Salomon & Gardner, 1984; Clark, 1983; Kearsley, Hunter & Sidel, 1983; Kulik, Bangert & Williams, 1983; Reiser & Gagne, 1982; Sheingold, Kane & Enderweit, 1983); Stimmel, Common, McCaskill and Durrett, 1981). While the influence of CBI research and evaluation data on CBI adoption for instruction is largely unknown, it is increasingly common for school districts to support local and costly CBI evaluation's (e.g. Sheingold, Kane &

Enderweit, 1983). Approximately 10% of all paper presentations at this convention in 1984 were devoted to this issue (Clark, 1984). Five years ago, Kulik and his colleagues located more than 500 studies in a search for classroom-based comparisons of CBI and conventional teaching (Kulik, Kulik & Cohen, 1980). It is relatively safe to guess that the number of locally supported research and evaluation studies is also larger today.

Meta-analytic Studies. The most systematic attempts to review CBI research, the meta-analytical surveys (Glass, 1976), report average gains of .5 standard deviations in examination scores over other instructional strategies (Clark, In press). This averages to as much as a fifteen point advantage for CBI students on a 100 point final examination. These findings are consistent for all levels of education including college (Kulik, Kulik & Cohen, 1980), secondary schools (Kulik, Bangert & Williams, 1983), and elementary schools (Kulik, Kulik & Bangert-Drowns, 1984). Yet, there have been recent and serious disputes about the validity of the CBI studies.

Disputes: The "Same Teacher" Effect. Clark (1983, 1984, In press), for example, has claimed that most of the research that forms the basis of CBI reviews may be seriously confounded. He argues that CBI contributions to science learning gains are susceptible to compelling rival hypotheses. The source of the rival explanations, Clark claimed, was the uncontrolled but systematic influence of different instructional methods, curriculum content and/or novelty. In support of the claim, Clark (In press) points to evidence that college CBI effect sizes reduce to insignificant levels when the same teacher designs both the CBI and the traditional treatments. He suggested that when only one designer produces both treatments, the opportunity for controlling all but the CBI aspects of the main treatment are greatest. In the "same teacher" studies therefore, the differences between CBI and traditional forms "wash out". He also points to evidence in both the college and secondary school CBI studies that the strongest effect sizes are in the shorter term studies which suggests a novelty effect. If this research is confounded as Clark claims, the expensive adoption of computers by schools in order to provide compensatory enhancement of science learning may require more consideration.

Alternatives: The "John Henry" Effect. There are, of course, other plausible explanations for the "same teacher" findings and for the stronger effects in shorter duration studies than those advocated by Clark. It is possible that when the same teacher designs both CBI and "traditional" (usually classroom lecture) presentations, a "John Henry"or compensatory rivalry effect is the result (Heinich, 1970). Here, the teacher consciously or unconsciously reduces the learning impact of the new technology. Heinich (1984) has recently reviewed studies where teachers and other workers engage in such rivalry when their jobs are threatened. If this phenomenon is operating in these studies,

there would be indications that the "same teacher" is producing a weak CBI presentation while enhancing the more compatible "traditional" treatment. If this is the case, the actual effect size of CBI treatments is even greater than the meta-analytical research reviews have suggested.

Resolving the Disputes. One generally suspects that disputes like the one surrounding CBI will be resolved by finding "middle distance" solutions -- in this case that some of the research is confounded and some is not. One obvious solution in this event would be to locate the unconfounded research and repeat the meta analysis. Yet, the problem here is that there is very little disagreement over whether the CBI research is confounded. The disagreement involves the nature of the confounding. If CBI effects are due to uncontrolled method, content and/or novelty then resources invested in that technology for science education are not going to have learning consequences. If the confounding is due to uncontrolled variables such as the "John Henry" effect a correct identification of the source will allow educators to more realistically assess the potential of CBI for education. This report describes a reexamination of the studies which have formed the basis of these meta-analytic reviews to determine the source and the extent of confounding. The discussion turns next to an examination of the evidence offered by both the "same teacher" and the "John Henry" advocates.

EVIDENCE FOR DIFFERENT TYPES OF CONFOUNDING

Evidence for the John Henry Effect. Advocates of the use of computers for instruction notice, for example that: "The ways in which the labor movement tried to protect its members from the encroachment of technology are very similar to how teacher groups seek to maintain the labor intensive character of instruction." (Heinich, 1984, p. 78). The basic argument here is that teachers and/or curriculum specialists engage in at least two activities that lead to research confounding. First, when they develop and/or administer comparison treatments they denegrate the treatment(s) that are threatening. This is the alternative hypothesis to the "same teacher" CBI effect suggested by Kulik (personal communication). Examples of this phenomenon are noted by Cooley & Lohnes (1976) in their analysis of their review of large-scale evaluations of instructional programs. The source of confounding was common enough for them to recommended that future evaluations include measures of "opportunity to learn" from new instructional approaches. The second source of confounding occurs when teachers in the non-CBI control or comparison groups invest extraordinary effort to "defeat" the threatening technology. This source of confounding was documented by Heinich (1970) in a series of studies of filmed and televised science courses such as the Wisconsin experiment on the Harvey White physics course. It is reasonable to suggest that the actual influence of CBI on learning cannot be accurately estimated in the

face of John Henry types of confounding.

Evidence for the Same Teacher Effect. Those who hold alternative views also find support in earlier research. Freeman (1924; Quoted in Saettler, 1964) challenged the simple "new versus old media" comparison study and recommended multivariate studies with more careful controls. Mielke (1968) described problems and solutions for the design of instructional television (ITV) research which are very relevant to the "same teacher" approach to CBI studies. He advised that when the learning benefits of newer media are being compared with older media, only those elements which are the fixed criterial or defining attributes of the media can be allowed in the comparison. This means that if, for example, computers are being compared with television or teachers as instructional tools, all treatments must contain exactly the same instructional methods to present the same curriculum content. If different methods are employed to teach the same content the results of such research would be confounded (i.e. we would not be able to distinguish the effects of method from those of medium on subsequent changes in learning scores). In fact, a moment's reflection will establish that under these conditions there may be no reason to conduct such a study in the first place. There is little reason to believe that when all non-medium variables have been controlled learning changes attributable to medium will occur.

Those who make the "same teacher" argument usually suggest that CBI research treatments employ very different levels of design effort and instructional method than contrasting or "traditional" treatments. The typical CBI treatment uses some form of drill and practice plus rule/example instructional method to teach. These CBI programs are often the result of considerable design and programming effort. Contrasting treatments, regardless of the medium, usually take less effort and involve different (and often vaguely described) methods such as "teacher lecture plus discussion". We cannot attribute learning differences found in these comparisons to "CBI". The "same teacher" evidence in the meta-analytic reviews is, in this view, only one indication of a much larger and more insidious confounding.

The Key Issues that separate these two approaches focus on both the sources of CBI research confounding and the impact on instruction from computers if the error were eliminated from future research. Those who point to compensatory rivalry interference in past studies expect that CBI effects are larger than current estimates. They generally support the view that eliminating these (and other) sources of error will allow us to emphasize computer attributes (e.g. Clark & Salomon, In press; Clark, 1983; Clark, In press; Papert, 1980; Salomon, 1979) These attributes are expected to play a key role in the future design of CBI by cultivating unique cognitive skills important for learning gains.

The alternative view points to evidence that there is no reason to expect unique computer effects on learning in the first place. They are confident that as in past studies, learning gains attributed to CBI are a consequence of failures to control differences in instructional method and content. They expect even more evidence that the new computer attribute variables actually belong to "instructional method" constructs. They point to evidence that attribute variables are not unique products of computers. They are only convenient operational forms of "methods" which might be delivered by many media with the same learning consequences. This point of view implicitly argues for a splitting of future CBI research questions into two different types: 1) questions dealing with the consequences of delivering and assessing science instruction via CBI versus "other" media; and 2) questions dealing with those facets of instruction which influence learning gains. Delivery questions involve factors which influence instructional outcomes such as cost, student access (e.g. availability, equity) and attitude (e.g. motivation), reliability, and organizational changes. In this view, learning is not influenced by the delivery vehicle or any unique aspect of that vehicle but by the interaction of instructional tasks, student aptitudes and instructional method (e.g. Clark, 1983).

Generally then, researchers with different views about the source of confounding tend to agree that; 1) CBI studies are confounded, 2) Identifying the source of the confounding would aid both the present interpretations of past research and the scientific and utilitarian applications of future studies. In order to set the stage for the specific objectives of the proposed study, the next section provides a brief review of the concept of confounding and describes its relevance in the context of research on the use of CBI in science instruction.

CONFOUNDING IN RESEARCH

In confounded research causes or effects cannot be unambiguously identified. When causes are confounded, it is still possible to determine the size of the effect of the treatment on the criterian (effect) measure. However, it is not possible to determine which of many uncontrolled parts of the treatment contributed to the difference. In instructional research, the confusion might involve an unintended mixing of a planned teaching method variable with a an unplanned but influential variable such as "experimenter expectancy".

Confounding does not damage our judgments about whether cause and effect are related (i.e. internal validity). Instead, it is a threat to "external validity" -- the generalizability of research findings. Its determination influences both the practical and theoretical application of research findings. In any analysis of research one generally begins with internal validity concerns and

then considers external validity. When the extent of internal validity has been determined one asks, "Given that the relationship here is plausibly causal and in the intended direction, what are the particular cause and effect constructs involved?". One of the critical elements necessary to answer this question is to insure that we can generalize from the operational definition of a construct to referent constructs in other studies or in practice.

Research On Instruction. Confounding is a common problem with all instructional research. In the past there has been an applied orientation to most studies in this area. Research is used to inform and guide instructional design. This requires that the statistical conclusions of research be accompanied by external validity. Meta analyses only assess internal validity. In a review of problems similar to those which confront CBI analyses, Shaver (1983) concluded that "Statistical significance is not a meaningful criterian for judging whether teaching methods were adequately controlled" (p. 8). Actually, the problem is common to all applied research.

By its nature, applied research concerns itself more with effects than with causes. Cook & Campbell in a useful discussion of the problem explain that "...what one wants to see is evidence that the social problem being addressed is at least partially ameliorated...Thus, great care goes into measuring outcomes...It is our distinct impression that most applied experimental research is much more oriented toward high construct validity of effects than of causes." (1979, p. 63). Yet, the attention to effects rather than causes is ultimately and paradoxically self defeating. Since we do not know what causes the measured change we cannot reliably repeat our successes or construct theory. The emphasis is on whether things worked rather than why. The result provides us with knowledge about the past applications but no useful predictions about ways to influence the problem in the future (Kerlinger, 1977).

Construct Validity Benefits. The volume of research on CBI applications offers a number of potential advantages for enhancing future theory development and practice. On the one hand there is the opportunity to increase its practical utility by determining whether the research is confounded with uncontrolled variables. While this is a negative activity, the results of such an analysis would contribute to the confidence level of those who purchase, program and employ CBI in science instruction. On the other hand, an analysis of the construct validity of existing studies would also lend information about what "may have" caused the impressive learning gains aside from CBI or computer attributes. Such information lends itself to the generation of hypotheses and theoretical models to guide future research. As Cook & Campbell (1979) suggest, "The important point is that construct validity consists of more than merely assessing the fit between planned constructs and the operations that were tailored to these

constructs. One can use the obtained pattern of data to edit one's thinking about both the cause and the effect constructs, and one can suggest, after the fact, other constructs that might fit the data better than those with which the experiments began." (p. 69). Construct validation of CBI research is the goal of this research.

Meta-Analytic Studies. Recently developed technology for " meta-analytic" reviews of existing studies (Glass, 1976; Slavin, 1984) are one method of gaining indicators of CBI confounding. These studies aggregate all acceptable research in an area and reduce their findings to a common metric. The purpose is to determine their "effect size" (the difference on a criterian measure between experimental and control groups, divided by the standard deviation of the control group). While the concern of these studies is internal validity, studies are coded by different design features. This allows us to estimate whether effect size might be related to design elements. In the case of the CBI meta-analyses, there are strong indications of confounding. One such instance was obvious in the Kulik, et al. (1980) meta-analysis of college CBI.

Table 1 presents selected correlations between effect size and different features of CBI studies from Kulik, et al. (1980). For example, as the extent of computer use for simulations increases in CBI studies, the learning "effect" tends to decrease (-.18). As the tendency for the same instructor to design both CBI and the traditional instructional treatments increases, the effect size also decreases (-.27). This latter correlation is the basis for competing claims (e.g. that the same instructor was designing the same teaching method into both CBI and control treatments which eliminated method as a source of confounding -- versus -- the same instructor was threatened by the CBI and purposely reduced its teaching effect).

Table 1

Correlations of Selected CBI Study Features and Achievement from Kulik, Kulik & Cohen, (1980, p. 536).

Study Characteristics	Correlation with Effect
Use of Computer for Simulation	-.18
Use of Computer for Programming.....	-.14
Extent of use.....	.15
Control for Same Instructor Effect.....	-.27*
Duration of Treatment.....	-.08
Content Emphasis on "hard" discipline.....	-.18
Content Emphasis on "life" studies.....	.16

* $p<.05$

Since previous thinking about CBI research indicated the possibility of instructional method and compensatory rivalry problems, these are some of the main issues to be explored further. Kulik's data clearly indicates that control problems exists in many of the studies selected for analysis. If the same instructor designation is a proxy for studies where method and content are uncontrolled, one would want to know which methods are responsible for the impressive achievement gains from these studies. If teachers are purposely reducing the effects of CBI because of perceived threats, those who design instruction will benefit from the knowledge. Method confounding provides an opportunity to misinterpret the "active ingredient" of CBI as the computer when it is the choice of "instructional method" which influences learning gains. Unidentified compensatory rivalry problems lead to the assumption that CBI is less effective. Method confounding limits the external validity (generalizability) of research whereas compensatory rivalry is a threat to internal validity. Since reports of CBI studies typically provide more extensive descriptions of research procedures, it is expected that they will give indications concerning the source of the confounding.

The major reason for selecting studies from meta-analytic reviews to determine construct validity is that systematic prior effort has been made to cull out studies which lack internal validity controls. Internal validity is a necessary (but not sufficient) prior condition for external validity. All too often, "box score" talleys of studies provided in research reviews are selected according to the reviewers bias (Slavin, 1984). In the meta-analyses performed by Kulik and his colleagues (Kulik, Kulik & Cohen, 1980), Kulik Bangert & Willams, 1983 and Kulik, Kulik & Bangert-Drowns, 1984) all available studies were surveyed and those analyzed were selected when: 1) They were methodologically sound (e.g. random subject selection, no teaching to the test in one treatment, no attrition problems); 2) Studies measured outcomes of both CBI and control groups; and 3) The treatments were presented in a field (i.e. classroom/school) location. The results of these analyses indicate that the treatments used in these studies produced a robust and educationally significant gain in achievement. Therefore, these studies would be the first place to look for a determination of the construct(s) which caused the gains. So, the discussion turns next to the objectives, method and products of the proposed research.

Objectives of the Proposed Research The study had three primary objectives:

1) To secure and review a representative sample of the studies that formed the basis of the meta-analyses of CBI influence on achievement.

2) To determine the type and extent of either method or compensatory rivalry confounding and to note other design problems with these studies if they exist.

3) To summarize the results of the construct validation in two forms:

   a) Suggestions directed to those who develop and/or use CBI in education.

   b) Suggestions directed to those who conduct research on CBI applications to teaching.

METHOD

Identification and Selection of Studies. Researchers selected a 30% random sample of the 128+ studies utilized by Kulik and his colleagues for their meta-analyses of college (Kulik, J., Kulik, C., and Cohen, 1980), Secondary school (Kulik, J., Bangert, and Williams, 1983) and Elementary school studies (Kulik,J, Kulik, C, and Bangert-Drowns, 1984).

Coding the Design Features of Studies. All studies were read by two experienced researchers who utilized specific forms (see Appendix A) for coding the manifest features of each study of interest.

Features to be Coded. The determination of confounding in research involves a two stage process. First one must determine which variables were operationalized in the treatments. This determination involves the listing of the critical defining features of each construct. Second, it is necessary to isolate and identify "cognate constructs" (Cook & Campbell, 1979). The first step is an attempt to determine whether some commonly understood and accepted version of "CBI" was manipulated. The second step involves a check to see if common (or uncommon) sources of threats to external validity were plausibly involved and active in the study. The only difficulty expected with the first determination is that more recent studies may seek to investigate CBI "attributes" rather than gross comparisons between CBI and "traditional" instruction. It is the second determination, the identification of plausible rival hypotheses, that requires the most intensive effort. coding sheets were developed to allow reviewers to determine the robustness of many plausible external validity threats including: 1. Surplus construct irrelevancies (e.g. standardizing of method and content between treatments); 2. Inadequate preoperational explication of constructs (e.g. was the CBI treatment based on what is commonly understood by that term?) 3. Mono operational or mono method bias (i.e. is more than one operational procedure examined?); 4.

Experimenter expectancies (i.e. did the researcher deliver one or more treatments?); 5. A confounding of construct levels (i.e. does one treatment exhibit more intensity or provide more opportunity to learn than another); 6. Compensatory rivalry (i.e. is there evidence that control teachers competed with the CBI treatment?)

RESULTS AND DISCUSSION

Characteristics of the Study Sample. We randomly selected 42 of 128 studies and each of 2 reviewers independently coded 40 features of each study. The authors served as reviewers and agreed on 83% of the 40 items coded on each study. It was determined that 2 studies (26 and 44, Study numbers correspond to list at end of paper) were uninterpretable and could not have been used in the meta-analyses (reports only provided mean differences between groups) so the final sample reported here totaled 40 studies. Table 2 presents the breakdown of the studies by source (dissertation, journal article or special report) and level (primary, secondary or college) and by expected and obtained frequencies. Since the expected frequency in each source/level cell matched the obtained frequency, we presumed that the sample was representative. We coded each study separately even though one report might give details of up to three studies (e.g. 20, 34 & 40).

. . . . . .

INSERT TABLE 2 ABOUT HERE

. . . . . .

These studies tended to be a bit dated with a mean publication year of 1973 and an SD of 3.7 years. The number of subjects participating varied from 18 to 5,000 but over 75% of the studies utilized more than 50 participants. A majority of the studies taught either mathematics (n = 18) or "hard" science (physics, chemistry, biology; n = 9) content. Over 70% of the studies utilized treatments which lasted over 10 hours with a time range of from 1 to 60 hours reported and a mean treatment time of 18.7 hours (SD = 13.6).

Study Design Quality. The reviewers agreed that 10 studies in the sample lacked any serious design problems (12, 13, 17, 18, 20, 21, 22, 23, 27, and 37). In fact, study number 17 reported by Green and Mink (1972) and conducted by the psychology faculty at Macalester College in Minnesota could easily serve as a model for this type of investigation. The two, one-year studies (20, 21) reported in the dissertation by Hatfield (1969) at University of Minnesota are outstanding examples of the way careful research in this area evolves.

The Reviewers also identified a great number of studies which had very obvious design problems. For example, of the 40 studies which were interpretable, 18 provided CBI to subjects

which was apparently not matched by non-computer delivered, control group instruction. In 3 other studies (19, 34, 35) the computer group was forced to persist until an achievement criterian had been reached while subjects in the control treatments were allowed to quit instruction whenever they wished. In two other studies (1, 7) there was obvious disproportional subject attrition from the CBI group (perhaps due to treatment difficulty). At least 2 studies (8, 14) had obvious selection problems that may have resulted in more able experimental than control groups. In only 10 studies of the 40 were subjects randomly assigned to treatments.

Perhaps the most serious design problem in over half of these studies is the providing of additional CBI which is not matched by non-computer control group instruction. One could hardly conclude from these studies that achievement advantages for the CBI group could be attributed to the computer. Instead, most of these studies should be categorized as "management" or "efficiency" investigations. In most cases the researchers were concerned about whether the computer could efficiently and non-reactively take over the delivery of drill and practice thus freeing the classroom teacher. One of these studies (12) was particularly well designed. Yet, none were examining the relative learning benefits of CBI versus traditional instruction.

Confounding of Instructional Method. The most common form of CBI instructional method (n = 15 studies) was some form of "drill and practice" similar to that made popular by Suppes & Morningstar (1969). This was the method most often provided CBI subjects in those studies with no "competing" control treatment (i.e. in 9, 12, 15, 29, 38, 39, 40, 41, 42). This control problem makes it impossible to conclude that the measured achievement gains were due to any aspect of the "computer" which was used to deliver the drill and practice. Achievement could have (and does) result from drill and practice delivered by nearly any medium including teachers. This point was made very dramatically in a series of CBI studies by Suppes and Morningstar (1969). They used computers to deliver additional drill and practice in mathematics to elementary students. Both experimental and control groups got the same instructional methods and content but the CBI group received an additional 5 to 8 minutes per day of computerized drill. Of course they outperformed the control groups in all but one instance. In that school the control group achievement was the highest of any group in this very large and well-managed study. Upon investigation Suppes found that in one of the control schools the principal had become alarmed at the pretest mathematics score levels of his students and had, on his own initiative, ordered 25 minutes of drill and practice in mathematics a day -- without the computer. We concur totally with Suppes & Morningstars' conclusion that the computer was not the important ingredient in the achievement gains. What was important was that "...the computerized drill and practice program required no extra time from the teacher and the student lost only 5 to 8 minutes from

nonmathematical subjects." (p. 346). Media may influence efficiency gains but instructional methods influence achievement gains.

The Same-Instructor Effect. There is considerable evidence in these studies (see Table 3) that when the same teacher taught both experimental (CBI) and control treatments (n = 15 studies), the chances for control of both instructional methods and curriculum content were significantly greater (Chi square value = 3.55, df 1,39, p.> .05). It was also obvious that method and content were not controlled in all of the "same instructor" studies which may be why the average effect size for the CBI treatments do not reduce to zero when these studies are separately considered in meta-analyses.

. . . . . .

INSERT TABLE 3 ABOUT HERE

. . . . . .

Even more important was the reviewers report that adequate method and content control produced significantly fewer outcomes which favored CBI (Chi Square = 4.95, df 1,39) p. > .05). In fact, in only 2 of 15 studies (12, 21) where method and content were controlled where there significant effects for CBI. It is interesting to note that both of these well controlled studies found differences favoring CBI only after reanalyses or delayed postests focused on the performance of students at different ability levels. In each case (Hatfield, 1960 and Lang, 1975) high ability students profited from computer programming of math or physics principles whereas low ability students had their learning depressed by the computer programming activity. In Hatfield (1960) the non-CBI treatment required performance of the same activity without the computer. In Lang (1975) the CBI instruction was an additional drill and practice exercise not matched by control group instruction. The type of task used by Hatfield requires a systematic procedural plan for using a principle to solve a problem (e.g. setting up a "lab exercise" to practice a physics principle). The plan can be developed without using a computer. In the Hatfield study, the lower ability students presumably had difficulty mastering the computer and the procedural planning. The computer programming task may have given the higher ability students the opportunity to employ their skills more ideosyncratically which seems generally to result in greater learning for them (e.g. Cronbach & Snow, 1979). However, even here the best hypothesis for the advantage of the high aptitude students may be in the fact that they were challenged by the difficulty of the CBI learning task, not by the computer per se (cf. the discussion by Clark & Salomon, In press).

. . . . . .

INSERT TABLE 4 ABOUT HERE

. . . . . .

Evidence for Compensatory Rivalry. It is difficult to

determine whether control group teachers exerted extraordinary effort to compete with the "new" computer instruction. However, there were 10 studies in this sample where significant differences were found favoring the control groups (2, 3, 4, 11, 12, 22, 33, 40, 41, 42). This is a fairly extraordinary finding in its own right but it also seems a promising place to look for the "John Henry Effect".

Of the 10 studies were control group advantages were found, 3 (2, 4, 33) were "same teacher" studies. We assume that rivalry would be unlikely when the same teacher designs and administers the computer treatment. In two of these studies (4, 33) significant effects were also found for CBI treatments and so we eliminated all three from consideration. In 2 of the remaining 7 studies (12, 22), reviewers felt that both the content and methods were controlled between groups and so they were eliminated from consideration. In 1 of the remaining studies (11) the computer was used only to test students and no instruction was provided. There was no indication of rivalry here. Another of the remaining studies (3) there was clear evidence that the subjects simply found the CBI treatment too difficult in the time allotted.

Of the three remaining studies (40, 41, 42) all were reported by Suppes & Morningstar (1969). In these studies matched control schools only had pre and post test measures taken. It is possible that in the control schools the pretest sensitized teachers to greater efforts and/or that there was contaminating information that control schools were competing with "computer" schools. The example (discussed above) one control group school principal instigated additional, non-computer drill after viewing the pretest math scores for his school is a compelling instance. Yet, three studies out of 40 do not provide strong evidence for a "John Henry Effect" interpretation of the meta-analytic reviews.

SUMMARY AND CONCLUSIONS

If the studies reviewed for this report are representative of the set represented in the meta-analyses conducted by Kulik and his colleagues, a number of tentative conclusions can be drawn:

1) Only about 25% of these studies lack serious design flaws.

2) In over 50% of the studies, there was a failure to control the amount ofinstruction received in the experimental and control groups. CBI groups received more instruction which probably was a major factor in the effect size estimates of the meta-analyses.

3) In about 40% of the studies the same teacher taught both CBI and control groups. These "same teacher" studies had significantly better teaching method and content controls than other studies.

4) In only 2 of the 15 studies where method and content were controlled were there significant differences favoring the CBI groups. In both of these studies significant differences were found only after delayed analyses examined the performance of students at different ability levels. This finding implies that only 5% of the meta-analytic studies have important achievement data favoring CBI instruction. Even here there are rival hypotheses that suggest the achievement differences are due to student trait and task interactions, not to the "computer".

5) There is no compelling evidence for the John Henry Effect (compensatory rivalry). One study clearly indicated that a control school had been sensitized by pretest scores and after extra non-computer drill and practice exercises, outperformed the CBI schools.

Suggestions for Research. Researchers who design CBI studies are urged to avoid gross comparisons of the relative achievement available from CBI and "traditional instruction". There is ample evidence in existing studies that no theoretical reason exists to ask such a question. Whenever adequately designed studies have asked the question, no differences in achievement have been found that may be unambiguously attributed to "computers". Instead, researchers are encouraged to investigate the relative efficiency and cost of CBI versus traditional instruction. It is possible that computers and other new technologies for delivering information (e.g. video disk) will significantly influence the cost, management, availability and/or the efficiency of instructional delivery to diverse student populations. In these efficiency studies, researchers are cautioned to control instructional method and content.

Suggestions for Practice. Practitioners interested in CBI are cautioned that the achievement gains attributed to the computer mode of delivery are probably due instead to the instructional methods employed in their "software". These methods could probably be delivered by a teacher. The contribution of the computer (and other new delivery technologies) may be to make instruction more efficient (free the teacher, reach more students with less per-student cost). Educators are urged to avoid rationalizing computer purchases by referencing the achievement gains in the meta-analytic studies. Computers make no more contribution to learning than the truck which delivers groceries to the market contributes to improved nutrition in a community. Purchasing a truck will not improve nutrition just as purchasing a computer will not improve student achievement. Nutrition gains come from getting the correct "groceries" to the people who need them. Similarly, achievement gains result from matching the correct teaching method to the student who needs it.

REFERENCES

Clark, R.E., (1983), Reconsidering research on learning from media. Review of Educational Research, 53(4), 445-460.

Clark, R.E., (April 1984) Learning from computers:Theoretical problems. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Clark, R.E., (1985). Confounding in educational computing research. In press for the Journal of Educational Computing Research, 1(2).

Clark, R.E., and Salomon, G. (1985). Media in Teaching, In press for Wittrock, M. (Ed.), Handbook of Research on Teaching. Volume III. New York: Macmillan.

Cook, T.D., & Campbell, D.T. (1979), Quasi-experimentation, Chicago:Rand McNally.

Cooley, W.W. & Lohnes, P. R., (1976) Evaluation research in education, New York: Irvington.

Cooper,16H.M., (1982), Scientific guidelines for conducting integrative research reviews, Review of Educational Research, 52(2), 291-302.

Cronbach, L.J. & Snow, R.E. (1979). Aptitudes and instructional method. New York: Irvington.

Glass, G.V. (1976), Primary, secondary and meta-analysis of research, Review of Educational Research, 5, 3-8.

Hatfield, L.L.(1969) Computer assisted mathematics:An investigation of the effectiveness of the computer used as a tool to learn mathematics. Unpublished doctoral dissertation from the University of Minnesota, University Microfilms No. 70-5569.

Heinich, R., (1970) Technology and the Management of Instruction, Washington D.C.:Association for Educational Communication and Technology.

Heinich, R., (1984), The proper study of instructional technology, Educational Communication and Technology Journal, 32(2), 67-87.

Jamison, D., Suppes, P., & Welles, S., (1974). The effectiveness of alternative instructional media:A survey. Review of Educational Research, 44, 1-68.

Kearsley, G. Hunter, B. & Sidel, R.J., (1983) Two decades of computer based instruction: What have we learned? T.H.E. Journal, 10, 88-96.

Kerlinger, F.N. (September, 1977), The influence of research on educational practice, Educational Researcher, 5-12.

Kulik, C., Kulik, J., and Bangert-Drowns, R., (1984) Effects of computer based education on secondary school pupils, Paper presented at the annual meeting of the American Educational Research Association, New Orleans, La.

Kulik, J., Bangert, R. & Williams, G., (1983) Effects of computer-based teaching on secondary school students., Journal of Educational Psychology, 75(1), 19-26.

Kulik, J., Kulik, C., & Cohen, P. (1980), Effectiveness of computer-based college teaching: A meta-analysis of findings. Review of Educational Research, 50, 525-544.

Lang, C.R., (1976) Computer graphic simulations in high school physics. Unpublished doctoral dissertation, Kansas State University. University Microfilms No. 72-32, 153.

Merrill, D.M., (1983) Component display theory, In Reigeluth, C. (Ed.), Instructional Design Theories and Models, Hillsdale, N.J.:Lawrence Earlbaum Associates.

Mielke, K.W., (1968) Questioning the questions of ETV research., Educational Broadcasting Review, 2, 6-15.

Papert, S., (1980) Mindstorms:Children, computers and powerful ideas, New York:Basic Books.

Quality of Education Data, (1984), 1983-84 Microcomputer use in schools. Denver Colorado:Quality Education Data, Inc.

Reiser, R. & Gagne, R., (1982) Characteristics of media selection models. Review of Educational Research, 52(4), 499-512.

Riccobono, J.A., (1984) Availability, use and support of instructional media, 1982-83. Corporation for Public Broadcasting. Washington, D.C.: National Center for Educational Statistics.

Saettler, P.A. (1968) A history of instructional technology. New York:McGraw Hill.

Salomon, G. (1979), Interaction of media, cognition and learning., San Francisco: Jossey Bass.

Salomon, G. & Gardner, H., (1984) The computer as educator: Lessons from Television Research. Cambridge, Mass.: Harvard Press.

Sheingold, K., Kane, J.H., & Enderweit, M.E., (1983) Microcomputer use in schools: Developing a research agenda. Harvard Education Review, 53, 412-432.

Shaver, J.P. (October, 1983), The verification of independent variables in teaching methods research., Educational Researcher, 3-9.

Slavin, R.E. (October, 1984), Meta-analysis in education: How has it been used?, Educational Researcher, 6-16.

Stimmel, T., Common, J., McCaskill, B. & Durrett, H.J., (1981), Teacher resistance to computer assisted instruction., Behavior Research Methods and Instrumentation. 13, 128-130.

Suppes, P. & Morningstar, M., (October 1969). Computer-assisted instruction, Science, 343 - 350.

RANDOMLY SELECTED AND REVIEWED STUDIES*

* Numbers in the left margin correspond to those used in the text. There were no studies assigned to numbers 10 and 32.

01 Alderman, D.L. (1978) Evaluation of the TICCIT computer-assisted instructional system in the community college. (ETS PR 78-10), Princeton, N.J.: Educational Testing Service.

02 Axeen, M.E. (1971) Teaching library use to undergraduates -- comparison of computer-based instruction and the conventional lecture. Urbana Ill: University of Illinois. ERIC Document No ED 014 316.

03 Boyd, A.L. (1973) Computer aided mathematics instruction for low-achieving students. Dissertation abstracts international, 33, 553A. (University Microfilms No. 73-17, 131).

04 Broh, C.A. (1975) Achievement and attitude with computer related instruction: A field experiment. Paper presented at the annual meeting of the American Political Science Association, San Francisco. ERIC No. ED 110 399.

05 Cox, S. (1974) Computer assisted instruction and student performance in macroeconimic principles. The Journal of Economic Education, 6(1), 29 - 37.

06 Chiang, A. Stauffer, C. & Cannara, A. (1978). Demonstration of the use of computer-assisted instruction with handicapped children, Final Report. (Report No. 446-AH-60076A), Arlington,VA: RMC Research Corp. (ERIC NO. ED 166 913).

07 Confer, R.W. (1971). The effect of one style of computer

assisted instruction on the achievement of students who are repeating general mathematics. Dissertation Abstracts International, 32, 2399A, University Microfilms No. 71-29 729.

08 Coffman, W.E. and Olsen, S.A. (1980) The First Two Years of PLAN: An Evaluation of Program Impact. Iowa City Iowa: Iowa Testing Programs. ERIC No. ED 190 674.

09 Crawford, A.N. (1970), A pilot study of computer assisted drill and practice in seventh grade remedial mathematics. California Journal of Educational Research, 21, 170 - 174.

11 Davies, T.P. (1972). An evaluation of computer assisted instruction using a drill and practice program. Dissertation Abstracts International, 32, 6970B, University Microfilms No. 72-18 627.

12 Delon, F.G. (1970) A field test of computer assisted instruction in first grade mathematics. Educational Leadership., 28, 170 - 180.

13 Deloach, S.J. (1977) A comparative study of use of computer programming activities in an introductory college mathematics course for disadvantaged students. Unpublished doctoral dissertation, Indiana University, 1977. University Microfilms No. 78-5629.

14 Denton, J.J. (1973) A methodological study of a computer managed instructional program in high school physics. Dissertation Abstracts International, 33, 4966A, University Microfilms No. 73-7023.

15 Fejfar, F.L. (1969). ISU lab school fourth graders learn through CAI. Contemporary Education., 40, 296-297.

16 Gershman, J. & Sakamoto, E. (1981) Computer-assisted remediation and evaluation: A CAI project for Ontario secondary schools. Educational Technology, 21, 40-43.

17 Green, C. & Mink, W. (1973) Evaluation of computer simulation of experiments in teaching scientific methodology. ST.Paul, Minnesota.: Macalester College. ERIC No. 082 475.

18 Gallagher, P.D., (1970) An investigation of instructional treatments and learner characteristics in a computer managed instruction course. Unpublished doctoral dissertation, Florida State University, University Microfilms No. 71-7013.

19 Haberman, E.L., (1977) Effectiveness of computer assisted instruction with socially/emotionally disturbed children. Unpublished Doctoral Dissertation from the University of Pittsburgh, University Microfilms No. 77-21 221.

20 Hatfield, L.L. (1970) Computer assisted mathematics: An investigation of the effectiveness of the computer used as a tool to learn mathematics. Unpublished doctoral Dissertation from the University of Minnesota. First Study; conducted in 1969. University Microfilms No. 71-18 720.

21 Hatfield, L.L. (1970) Note: Same report as 20 but this is the second study, conducted in 1970.

22 Johnson, R.E. (1971), The effect of activity oriented lessons on the achievement and attitudes of seventh grade students in mathematics. Dissertation Abstracts International, 32, 304A. University Microfilms No. 73-3897.

23 Jones, J.E., (1974). Computer simulated experiments in high school physics and chemistry. Unpublished Doctoral Dissertation, University of Iowa. University Microfilms No 71-18, 720.

24 Jones, L.A. & Sorlie, W.E. (1976). Increasing medical student performance with an interactive, computer assisted appraisal system. Journal of Computer Based Instruction, 2(3), 57-62.

25 Kelly, A. C. TIPS and technical change in classroom instruction. American Economic Review, 62, 422-428.

26 Krumhout, O.M., Edwards, S. & Schwarz, G.A. (1969) A computer guided general education physics course. American Journal of Physics, 37(10), 995-1001.

27 Lang, C.R. (1976). Computer graphic simulations in high school physics. Unpublished doctoral dissertation from Kansas State University. University Microfilms No. 68-17 690.

28 Lee, A.L. (1974). A comparison of computer assisted instruction and traditional laboratory instruction in an undergraduate geology course. Unpublished doctoral dissertation from the University of Texas. University Microfilms No. 73-26 036.

29 Litman, G. (1977). Relation between computer-assisted instruction and reading achievement among fourth, fifth and sixth grade students. Unpublished doctoral dissertation from Northern Illinois University, University Microfilms No. 77-20 883.

30 Mandelbaum, J. (1974). A study of the effects on achievement and attitude, of the use of the computer as a problem solving tool with low performing tenth grade students. Unpublished doctoral dissertation from Temple University. University Microfilms No. 74-1809.

31 McEwen, N. & Robinson, A. (1976). Computer assisted instruction in secondary school french. Final Report. Edmonton: Alberta University. ERIC No ED 150 846.

33 Paden, J.S. (1970). An experimental study of the individualized instruction in high school physics using the computer to prescribe activities as a function of selected idiographic factors. Unpublished doctoral dissertation from the University of Missouri at Columbia, University Microfilms No. 71-8373.

34 Paden, D. W., Dalgaard, B.R. & Barr, M.D. (1977) A decade of computer-assisted instruction. The Journal of Economic Education. 9(1), 14-20. Note: Study 1, 178 Subjects.

35 Paden, D.W. et al. Note: Same as 34 above, Study 2, 382 Subjects.

36 Paden, D.W. et al. Note same as 34 above, Study 3, 63 subjects.

37 Skavril, R.V., (1974) Computer based instruction of introductory statistics. Journal of Computer Based Instruction, 1(1), 32-40.

38 Skavaril, R.V., Birky, C. Jr., Dahrkopf, R.E. & Knight, J.A. (1976) The use of CAI to provide problems for students in introductory genetics. Journal of Computer Based Instruction. 3(1), 13-20.

39 Smith, E.S. (1980), The effect of computer-assisted instruction on academic achievement, school daily attendance and school library usage at Margaret Murry Washington Career Center. Unpublished Doctoral Dissertation from George Washington University. University Microfilms No. 80-26 342.

40 Suppes, P. & Morningstar, M. Computer-assisted instruction. Science, 166, 343-350. Note: 1966-67 study with 880 subjects.

41 Suppes, P. & Morningstar, M. Note: same as 40 above, 1967-68 study with 1800 subjects.

42 Suppes, P. & Morningstar, M. Note: same report as 40 above, 1967-68 Mississippi study with 1900 subjects.

43 Weiss, E. (1971) An investigation of the relative effectiveness of two techniques of teaching a transitional course in physics on the college level. Unpublished doctoral dissertation form New York University. University Microfilms No. 71-28 569.

44 Wilson, P.M., (1982) Computer based education and proficiency testing: A model of cost-effectiveness. Paper presented at the July SALT Conference on training effectiveness and evaluation.

Appendix A

INFORMATION TO BE CODED CONCERNING MEDIA STUDIES (Clark & Leonard)

1. Author(s) /year_____
2. Academic level of Ss (yr 1 to 20+)_____
3. Total # of Ss _____
4. Ss Type: Norm___ Remedial____ LH_____ EH____ Other_____
5. General Subject matter (specify)_____
6. Merrill Content/Perf. type (1 to 10)_____
7. Ss selection: Random to tmt___ Self___ by class___ Other____
8. Duration of treatments: Hours_____ Days_____
9. # of groups in comparison: Exp_____ Control_____
10. Total # of statistical comparisons _____
11. # of Comparisons: NSD_____ Favor CBI_____ Favor Control ____
12. Same teacher both E and C groups? yes___ no ___ ?yes__?no__
13. Experimenter is instructor? one group___ all groups___
14. Content controlled in E and C groups? yes___no___?yes__?no___
15. Method controlled in E and C groups? yes___no___?yes__?no___
16. CBI Method(s) used: (1 to 12)_____Other_____
17. Control method(s) used: (1 to 12)_____Other_____
18. Pretest given? yes_____no_____
19. Type of pretest? Teachers'___ Standardized(specify)_____
20. Statistical tests used for E/C comparison_____
    _____
21. Data used for tests: Raw scores___Adjusted ____(specify basis for adjustment)_____

Remarks (interesting features not included in coded items, e.g. compelling rival hypotheses, unexpected findings):

AERA STUDY DATA FILE: GUIDE TO COL DESIGNATIONS

Item	col's	GUIDE TO COL MEANING
1	1,2	Study # 01-44
2	3	academic level 1 ele 2 secondary 3 college 4 grad
3	4,5,6,7	total number of subjects
4	8	Ss type 1 norm 2 remedial 3 LH 4 EH 5 lo ses
5	9	1 math 2 phys/chem/bio 3 soc sci/hist 4 econ 5 read 6 lib sci 7 med 8 lang 9 geology
6	10	type of report: 1 dissertation 2 eval 3 journal
7	11	1 random 2 self 3 by class 4 matched 5 teacher
8	12,13	total number of hours in treatment (if E><C then E)
9	14,15 16,17	number of experimental groups number of control groups
10	18,19	total number of comparisons between E and C
11	20,21 22,23 24,25	number of comparisons NSD number of comparisons favoring Experimental group(s) " " " " Control "
12	26	Same teacher in E & C groups? 1 yes 2 no 3 NA
13	27	Experimentor is teacher? 1 yes 2 no 3 NA
14	28	Is instructional content conrolled? 1 yes 2 no
15	29	Is instructional method controlled? 1 yes 2 no
16	30,31	CBI method: 1 "traditional" or ? 2 Drill & Pract. 3 simul & model 4 write CAI prog 5 use #4 CAI prog. 6 systematically varied 7 test only 8 lab exercise 9 forced review 10 nothing/no treatment
17	32,33	Same as above in #16 for control group method
18	34	Pretest given? 1 yes 2 no 3 NA
19	35	Type of pretest: 1 teacher 2 standardized 3 na
20	36	Statistical tests used: 1 T 2 ANOVA/ANCOVA 3 REG
21	37	Data used for comparison tests: 1 raw 2 adjusted

22	38	Was CBI extra? 1 yes 2 no
23	39	1 Computer managed instruction 2 CBI
24	40	Obvious design problems:1 attrition 2 more e gp inst 3 selection 4 one tmt too difficult 5 control method previously inadequate 6 e to criterian but not c 9 no obvious design problems

Table 2

Number of Observed and Expected Frequencies of Studies by Publication Type and Grade Level

Type	Grade						Total	
	Elementary		Secondary		College			
Dissertation	4	(3)	9	(8.8)	4	(5.5)	17	(17.3)
Special Report	2	(3)	2	(1.5)	4	(4.8)	8	(9.3)
Journal	5	(5.8)	3	(2.8)	7	(6.4)	15	(15)
Total	11	(11.8)	14	(13.1)	15	(16.7)	40	(41.6)

Note: Expected study frequencies are in parentheses. Expected frequencies are the percent of the total studies in each grade/type cell.

Table 3

Chi Square Analysis of Teacher And Method/Content Control Features of Study Sample

		Same Teacher?	
		Yes	No
Both Method and Content Controlled?	Yes	8	6
	No	7	19

Table 4

Chi Square Analysis of Statistical Difference Results of Group Comparisons by Method/Content Control Features in Study Sample.

		Statistical Differences	
		Favor CBI	NSD or Favor Control
Both Method and Content Controlled?	Yes	2	12
	No	13	13